

#2
LDT
2601IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re U.S. Patent Application)
)
)
 Applicant:)
)
 Serial No.)
)
 Filed: November 29, 2000)
)
 For: INFORMATION RETRIEVAL)
 SYSTEM AND A COMPUTER)
 PRODUCT)
)
 Art Unit:)

*I hereby certify that this paper is being deposited
 with the United States Postal Service as EXPRESS
 mail in an envelope addressed to: Assistant
 Commissioner for Patents, Washington, D.C. 20231,
 on November 29, 2000.
 Express Label No.: EL769180853US
 Signature: A.F.*

CLAIM FOR PRIORITY

Assistant Commissioner for Patents
 Washington, DC 20231

Sir:

Applicant claims foreign priority benefits under 35 U.S.C. § 119 on the basis
 of the foreign application identified below:

Japanese Patent Application No. 2000-031228, filed February 8, 2000.

A certified copy of the priority document is enclosed.

Respectfully submitted,

GREER, BURNS & CRAIN, LTD.

By:
 James K. Folker
 Reg. No. 37,538

November 29, 2000
 300 South Wacker Drive
 Suite 2500
 Chicago, IL 60606
 (312) 360-0080

日本国特許庁

PATENT OFFICE
JAPANESE GOVERNMENT

PAJA-00030-12
1924.11.85
(312) 2.00 000.0



別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日
Date of Application: 2000年 2月 8日

出願番号
Application Number: 特願2000-031228

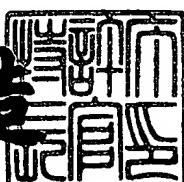
出願人
Applicant(s): 富士通株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

2000年 8月 18日

特許庁長官
Commissioner,
Patent Office

及川耕造



出証番号 出証特2000-3065999

【書類名】 特許願
【整理番号】 9951551
【提出日】 平成12年 2月 8日
【あて先】 特許庁長官殿
【国際特許分類】 G05F 15/40
【発明の名称】 文書情報検索装置および文書情報検索プログラムを記録
したコンピュータ読み取り可能な記録媒体
【請求項の数】 5
【発明者】
【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通
株式会社内
【氏名】 関根 秀憲
【特許出願人】
【識別番号】 000005223
【氏名又は名称】 富士通株式会社
【代理人】
【識別番号】 100089118
【弁理士】
【氏名又は名称】 酒井 宏明
【手数料の表示】
【予納台帳番号】 036711
【納付金額】 21,000円
【提出物件の目録】
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1
【包括委任状番号】 9717671
【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 文書情報検索装置および文書情報検索プログラムを記録したコンピュータ読み取り可能な記録媒体

【特許請求の範囲】

【請求項1】 ネットワーク上に存在する複数の情報を検索する情報検索装置において、

ユーザにより選択された情報に関する位置情報を位置情報データベースとして格納する格納手段と、

前記位置情報データベース内のそれぞれの位置情報の利用頻度を解析する解析手段と、

しきい値に対応した利用頻度の位置情報を収集基点とする所定の収集範囲に存在する情報を検索情報データベースとして収集する収集手段と、

ユーザにより指定された検索条件に基づいて、前記検索情報データベースから所望の情報を検索する検索手段と、

を備えることを特徴とする情報検索装置。

【請求項2】 前記解析手段は、前記利用頻度に対応する前記収集範囲を位置情報毎に算出することを特徴とする請求項1に記載の情報検索装置。

【請求項3】 前記解析手段は、前記利用頻度に対応する段階的な前記収集範囲を位置情報毎に算出することを特徴とする請求項1に記載の情報検索装置。

【請求項4】 前記解析手段は、算出された収集範囲と該収集範囲に対応しつつ収集基点となる位置情報をからなる収集基点位置情報データベースを構築し、前記収集手段は、前記収集基点位置情報データベースに基づいて、しきい値に対応した利用頻度の位置情報を収集基点とする所定の収集範囲に存在する情報を検索情報データベースとして収集することを特徴とする請求項2または3に記載の情報検索装置。

【請求項5】 ネットワーク上に存在する複数の情報を検索するための情報検索プログラムを記録したコンピュータ読み取り可能な記録媒体であって、

ユーザにより選択された情報に関する位置情報を位置情報データベースとして格納させる格納工程と、

前記位置情報データベース内のそれぞれの位置情報の利用頻度を解析させる解析工程と、

しきい値に対応した利用頻度の位置情報を収集基点とする所定の収集範囲に存在する情報を検索情報データベースとして収集させる収集工程と、

ユーザにより指定された検索条件に基づいて、前記検索情報データベースから所望の情報を検索させる検索工程と、

をコンピュータに実行させるための情報検索プログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、たとえば、インターネット上に存在する情報を検索する情報検索装置および情報検索プログラムを記録したコンピュータ読み取り可能な記録媒体に関するものである。

【0002】

近年、インターネットの利用率が加速的に増加しており、かかる背景より、インターネット上に存在する、例えば、HTML (HyperText Markup Language) で記述された文書情報の量も膨大なものとなっている。一般には、膨大な文書情報の中から所望の文書情報を検索する場合には、キーワード検索方式を探る検索エンジンを備える情報検索装置が用いられる。この種の情報検索装置では、ある文書情報を収集基点として、収集基点の文書情報にリンクされた文書情報を順次収集し、これらを検索情報としてデータベース化している。実際に検索を行う場合には、キーワード方式により検索情報のデータベースから複数（または単数）の文書情報を検索し、これらを検索結果としている。

【0003】

しかしながら、従来の情報検索装置では、一定の収集条件（リンク数、文書数、文書サイズ等）に基づいて、画一的に収集基点から文書情報を順次収集しているため、多数のユーザが満足し得る検索結果に結びつく検索情報を得ることが難しく、結果的に、検索精度が悪いという問題があり、かかる問題を効果的に解決

し得る手段、方法が切望されている。

【0004】

【従来の技術】

インターネットでは、サーバに蓄積されている文書情報にアクセスする手段（通信プロトコル）と文書情報の名前とを指定する規格として、URL（Uniform Resource Locator）がある。文書情報とは、例えば、HTMLで記述された情報（コンテンツ）をいう。例えば、サーバに蓄積されている文書情報のファイルを指定する場合、URLは、[プロトコル名://サーバ名/ファイル名]と記述される。言い換えれば、URLは、インターネット上における文書情報の位置を示す情報であり、以下では、URLを文書位置情報と称する。

【0005】

また、文書情報には、リンク先の他の文書情報の文書位置情報が含まれている場合がある。このような文書情報間のリンク状態が複数リンクに亘っている場合には、収集基点となる文書情報から芋蔓式に複数の文書情報の収集が可能である。上述した従来の情報検索装置では、一定の収集条件（リンク数等）に基づいて、収集基点となる文書情報から順次リンク先の文書情報を一定リンク数（収集範囲）で収集し、これらを検索情報としてデータベース化する。上記収集範囲は、情報検索装置を用いた検索サービス会社により決められており、ユーザのニーズが反映されたものではない。

【0006】

また、情報検索装置では、ユーザにより指定されたキーワードをキーとして検索情報のデータベースからキーワードを含む複数（または単数）の文書情報を検索し、これらを検索結果としている。ユーザは、検索結果から所望の文書情報を閲覧する。

【0007】

【発明が解決しようとする課題】

ところで、前述したように、従来の情報検索装置では、インターネット上の文書情報を収集する際に検索サービス会社により画一的に決められた、収集基点からの収集範囲の文書情報を収集し、この収集結果に基づいて検索が実行されてい

る旨を述べた。

【0008】

しかしながら、従来の情報検索装置では、たとえ、ユーザの利用ニーズが高い文書情報であっても、該文書情報が収集範囲外に存在する場合には収集結果、ひいては検索結果から漏れてしまうという問題があった。さらに従来の情報検索装置では、収集基点に存在する文書情報がユーザにさほど利用されていない場合であっても、当該文書情報に関連し、収集範囲の複数の文書情報が利用度に関わらず、一律に収集されてしまうため、検索結果として不要な文書情報が多く含まれ、検索精度が低下するという問題があった。すなわち、従来の情報検索装置では、検索効率が悪いのである。

【0009】

本発明は、上記に鑑みてなされたもので、検索効率を向上させることができる情報検索装置および情報検索プログラムを記録したコンピュータ読み取り可能な記録媒体を提供することを目的とする。

【0010】

【課題を解決するための手段】

上記目的を達成するために、本発明は、ネットワーク上に存在する複数の情報を検索する情報検索装置において、ユーザにより選択された情報に関する位置情報を位置情報データベースとして格納する格納手段と、前記位置情報データベース内のそれぞれの位置情報の利用頻度を解析する解析手段と、しきい値に対応した利用頻度の位置情報を収集基点とする所定の収集範囲に存在する情報を検索情報データベースとして収集する収集手段と、ユーザにより指定された検索条件に基づいて、前記検索情報データベースから所望の情報を検索する検索手段とを備えることを特徴とする。

【0011】

この発明によれば、解析手段により位置情報データベース内のそれぞれの位置情報の利用頻度が解析されると、収集手段により、しきい値に対応した利用頻度の位置情報を収集基点とする収集範囲の情報を収集される。ここで、収集される位置情報は、ユーザの利用頻度が高い情報である。そして、ユーザにより検索条

件が指定されると、検索手段により、検索情報データベースから所望の情報が検索される。

【0012】

このように、この発明によれば、しきい値に対応したユーザの利用頻度の情報を収集基点とする所定範囲の情報を収集し、収集された情報を被検索対象とするようにしたので、検索結果に利用頻度が高い情報が含まれる割合を高めることができ、結果として検索効率を向上させることができる。

【0013】

【発明の実施の形態】

以下、図面を参照して本発明にかかる情報検索装置および情報検索プログラムを記録したコンピュータ読み取り可能な記録媒体の実施の形態1および2について詳細に説明する。

【0014】

(実施の形態1)

図1は、本発明にかかる実施の形態1の構成を示すブロック図である。この図において、文書格納装置100および110は、複数の文書情報を格納する装置であり、ネットワーク200（たとえば、インターネット）を介して、クライアント400や検索装置300にそれぞれアクセスされる。ここでいう文書情報とは、例えば、HTMLにより記述された情報（コンテンツ）をいう。なお、一実施の形態で検索対象となる情報は、文書情報に限られるものではなく、JPEGファイル形式、MPEGファイル形式等のあらゆる形式の電子的情報である。検索装置300は、文書格納装置100および110から文書情報を収集する収集処理と、クライアント400により指定されるキーワードに基づいて収集された文書情報からキーワードが含まれる文書情報を検索する検索処理と、検索された文書情報の中からユーザにより選択された文書情報の存在位置を示す文書位置情報に基づいて、ユーザのニーズを解析する解析処理とを実行する。

【0015】

検索装置300において、文書位置情報データベース310には、図3に示したように、クライアント400により選択された日付、および文書情報の存在位

置を示す文書位置情報が格納部（図示せず）により格納される。解析部320は、ユーザのニーズの観点から文書位置情報データベース310を解析し、解析結果を図4に示した収集基点位置情報データベース330に反映させる。この解析部320の動作の詳細については、後述する。

【0016】

図4に示した収集基点位置情報データベース330は、収集処理時の収集基点位置情報および収集範囲からなるものである。この収集基点は、図2に示したように、検索情報収集部340（図1参照）が文書情報を収集する際の開始位置であり、収集範囲ARは、たとえば、1～6という1リンク単位の範囲である。従って、収集範囲ARが1である場合には、収集基点に存在する文書情報D₀から1リンクの範囲に存在する文書情報D₁₁およびD₁₂が収集される。

【0017】

同様にして、収集範囲ARが5である場合には、文書情報D₀から5リンクの範囲に存在する文書情報D₁₁、D₁₂、D₂₁、D₂₂、D₂₃、D₃₁、D₃₂、D₃₃、D₄₁、D₄₂、D₅₁およびD₅₂がそれぞれ収集される。なお、この場合には、収集範囲AR=6に存在する文書情報D₆₁は、対象外として収集されない。

【0018】

図1に戻り、検索情報収集部340は、収集基点位置情報データベース330に基づいて、文書格納装置100および110から収集範囲の文書情報を収集し、これを検索情報データベース350に格納する。図5に示した検索情報データベース350は、「インデックス」、当該検索情報（文書情報）に含まれるキーワードに対応する「キーワード」、当該検索情報（文書情報）に関する文書位置情報に対応する「URL」、当該検索情報（文書情報）に含まれる本文の一部（文字列）に対応する「第1表示情報」および「第2表示情報」のフィールドからなる。

【0019】

図1に戻り、検索部360は、クライアント400からのキーワード（以下、入力キーワードという）をキーとして、検索情報データベース350（図5参照）を検索し、入力キーワードと一致するキーワードを含む文書情報の文書位置情

報（URL）、第1表示情報および第2表示情報を検索結果（検索結果画面G₂：図15参照）としてクライアント400へ送信する。

【0020】

つぎに、図1に示した解析部320の動作について図6に示したフローチャートを参照しつつ説明する。図1において、後述する検索処理が実行されると、検索部360からは、検索結果がクライアント400へ送信される。そして、ユーザーにより、検索結果から所望の文書情報が選択されると、クライアント400は、当該文書情報が蓄積されている、たとえば、文書格納装置100へアクセスし、当該文書情報をダウンロードする。

【0021】

また、クライアント400は、選択された日付および選択された文書情報に関する文書位置情報を検索装置300へ送信する。これらの日付および文書位置情報は、図3に示した文書位置情報データベース310に格納される。上述した動作が繰り返されることにより、文書位置情報データベース310には、ユーザが実際に選択した文書情報に関する日付および文書位置情報が順次格納される。

【0022】

ここで、一定時間間隔をおいて解析部320は、文書位置情報データベース310に基づく解析処理を実行する。すなわち、図6に示したステップSA1では、解析部320は、文書位置情報データベース310（図3参照）の文書位置情報毎に選択回数を集計し、集計結果を図8に示した一時格納テーブルT₁に格納する。この一時格納テーブルT₁において、たとえば、1レコード目の文書位置情報（<http://www.abcdefg.co.jp/hypertext/newinfo/>）は、ユーザにより5回選択されている。

【0023】

図6に示したステップSA2では、解析部320は、一時格納テーブルT₁の選択回数およびつぎの（1）式から文書位置情報毎の選択頻度を算出する。この選択頻度は、総選択回数に対する当該文書位置情報の選択回数のパーセンテージを表す。

$$\text{選択頻度} = ((\text{当該文書位置情報の選択回数}) / (\text{全文書位置情報の選択回})$$

数の合計)) × 100 . . . (1)

【0024】

つぎに、解析部320は、算出された各選択頻度を図9に示した一時格納テーブル T_2 に格納する。この一時格納テーブル T_2 において、たとえば、1レコード目の「頻度」(選択頻度)は、10%である。ステップSA3では、解析部320は、図9に示した一時格納テーブル T_2 に対して、「頻度」(選択頻度)をキーとして、降順となるようにソートをかけた後、それぞれの文書位置情報に優先順位を付ける。

【0025】

つぎに、解析部320は、ソート結果を図10に示した一時格納テーブル T_3 に格納する。ステップSA4では、解析部320は、図10に示した一時格納テーブル T_3 で「頻度」(選択頻度)がしきい値(たとえば、10%)未満のレコード(優先順位、文書位置情報、頻度)を削除する。ここで、しきい値は、つぎの(2)式から算出される。

$$\text{しきい値} = ((\text{選択頻度の最大値}) + (\text{選択頻度の最小値})) / 2 \quad \dots (2)$$

【0026】

同図に示した例では、3レコード目から10レコード目が削除される。この削除されたレコードは、ユーザのニーズが極めて低い、言い換えれば、ほとんど利用されない文書情報に関する文書位置情報である。ステップSA5では、解析部320は、つぎの(3)式に基づいて、選択頻度から収集範囲を算出し、これを図11に示した一時格納テーブル T_4 に格納する。

$$\text{収集範囲} = (\text{当該選択頻度} / \text{選択頻度の最大値}) \times \text{収集範囲の最大値} \quad \dots (3)$$

ただし、(3)式の算出結果において端数は切り上げる。図11に示した1レコード目の収集範囲は、5(収集範囲AR=5:図2参照)であり、2レコード目の収集範囲は、1(収集範囲AR=1:図2参照)である。

【0027】

ステップSA6では、解析部320は、一時格納テーブル T_4 の情報(文書位

置情報、収集範囲)を図4に示した収集基点位置情報データベース330に追加するという追加処理を実行する。すなわち、図7に示したステップSB1では、解析部320は、一時格納テーブルT₄内に未追加の文書位置情報があるか否かを判断し、この場合、判断結果を「Yes」とする。

【0028】

ステップSB2では、解析部320は、一時格納テーブルT₄内の未追加の文書位置情報が収集基点位置情報データベース330の「収集基点位置情報」にすでに存在しているか否かを判断する。この判断結果が「Yes」である場合、ステップSB4では、解析部320は、収集基点位置情報データベース330における、収集範囲を、図11に示した収集範囲に更新する。

【0029】

一方、ステップSB2の判断結果が「No」である場合、図11に示した文書位置情報を、図4に示した「収集基点位置情報」として収集基点位置情報データベース330に追加する。以後、ステップSB以降1では、上述した動作が繰り返される。そして、追加処理が終了すると、解析部320は、ステップSB1の判断結果を「No」とし、一連の解析処理を終了する。

【0030】

つぎに、図1に示した検索情報収集部340の動作について、図12および図13に示したフローチャートを参照しつつ説明する。図12に示したステップSC1では、検索情報収集部340は、収集基点位置情報データベース330(図4参照)の最初の収集基点位置情報(この場合、<http://www.is.abcdefg.co.jp/qa/qa1-10.html>)を取得する。ステップSC2では、検索情報収集部340は、収集基点位置情報データベース330からすべての収集基点位置情報を取得したか否かを判断し、この判断結果が「Yes」である場合、終了する。

【0031】

この場合、ステップSC2の判断結果が「No」であるものとすると、ステップSC3では、検索情報収集部340は、ステップSC2で取得した収集基点位置情報に基づいて、ネットワーク200を介して、図2に示した収集基点に存在する文書情報D₀を文書格納装置110から取得する。ステップSC4では、検

索情報収集部340は、取得した文書情報D₀におけるキーワード、URL、第1表示情報および第2表示情報を検索情報データベース350（図5参照）に追加する。

【0032】

ステップSC5では、検索情報収集部340は、当該文書情報D₀にリンク先の文書位置情報が含まれているか否かを判断する。この場合、図2に示したように、文書情報D₀には、リンク先の文書情報D₁₁およびD₁₂のそれぞれに関する文書位置情報が含まれているものとする。一例として、図16には、上記文書情報D₀をクライアント400の表示部（図示せず）に表示させた場合の文書画面G₃が図示されている。この文書画面G₃の表示領域Dには、文書情報D₀が表示されているとともに、文書情報D₀のリンク先の文書情報D₁₁およびD₁₂のそれぞれの文書位置情報が表示されている。

【0033】

従って、検索情報収集部340は、ステップSC5の判断結果を「Yes」とし、ステップSC7の処理を実行する。なお、ステップSC5の判断結果が「No」である場合、ステップSC6では、検索情報収集部340は、収集基点位置情報データベース330（図4参照）のつぎの収集基点位置情報を取得する。以後、ステップSC2以降の動作が繰り返される。

【0034】

この場合、ステップSC7では、検索情報収集部340は、リンク先の文書情報を収集するリンク先文書情報収集処理を実行する。すなわち、図13に示したステップSD1では、検索情報収集部340は、リンク先位置情報データベースを作成する。この場合、リンク先位置情報データベースは、図2に示した文書情報D₀に含まれるリンク先の文書位置情報（文書情報D₁₁の文書位置情報および文書情報D₁₂の文書位置情報）からなる。

【0035】

ステップSD2では、検索情報収集部340は、リンク先位置情報データベースから全てのリンク先の文書情報を取得したか否かを判断する。ここでいう全てのリンク先の文書情報とは、図4に示した1レコード目の収集範囲5（図2：収

集範囲 A R = 5) に存在するものをいう。この場合、判断結果を「No」とし、ステップ SD 3 では、検索情報収集部 340 は、リンク先位置情報データベースにおける最初のリンク先位置情報に対応する文書情報 D₁₁ (図 2 参照) を文書格納装置 110 から取得する。

【0036】

ステップ SD 4 では、検索情報収集部 340 は、取得した文書情報 D₁₁ を検索情報データベース 350 に追加する。ステップ SD 5 では、検索情報収集部 340 は、収集範囲の最後であるか否かを判断し、この判断結果が「Yes」の場合、ステップ SD 11 の処理を実行する。この場合、検索情報収集部 340 は、ステップ SD 5 の判断結果を「No」とする。

【0037】

ステップ SD 6 では、検索情報収集部 340 は、文書情報 D₁₁ にリンク先の文書位置情報が含まれているか否かを判断する。この場合、図 2 に示したように、文書情報 D₁₁ には、リンク先の文書情報 D₂₁ に関する文書位置情報が含まれているものとする。従って、検索情報収集部 340 は、ステップ SD 6 の判断結果を「Yes」とする。なお、ステップ SD 6 の判断結果が「No」である場合、ステップ SD 11 では、検索情報収集部 340 は、リンク先位置情報データベースからつぎのリンク先の文書位置情報を取得し、ステップ SD 2 以降の処理を繰り返す。

【0038】

この場合、ステップ SD 7 では、検索情報収集部 340 は、リンク先位置情報データベースを作成する。この場合、リンク先位置情報データベースは、図 2 に示した文書情報 D₁₁ に含まれるリンク先の文書位置情報 (文書情報 D₂₁ の文書位置情報) からなる。ステップ SD 8 では、検索情報収集部 340 は、全てのリンク先の文書情報を取得したか否かを判断し、この場合、判断結果を「No」とする。なお、この判断結果が「Yes」である場合、ステップ SD 6 以降の処理を繰り返す。

【0039】

この場合、ステップ SD 9 では、検索情報収集部 340 は、リンク先の文書情

報D₂₁を文書格納装置110から取得する。ステップSD10では、検索情報収集部340は、文書情報D₂₁を検索情報データベース350に追加する。以後、上述した動作が繰り返されることにより、図2に示した収集基点から収集範囲A R=5に存在する複数の文書情報が取得された後、検索情報データベース350に順次追加される。

【0040】

そして、ステップSD2の判断結果が「Yes」になると、検索情報収集部340は、図12に示したステップSC6の処理を実行する。そして、ステップSC2の判断結果が「Yes」になると、検索情報収集部340は、一連の収集処理を終了する。この状態では、検索情報データベース350には、ユーザによる利用頻度が高い文書情報のみが格納されている。

【0041】

つぎに、検索部360の検索処理について説明する。図1に示したクライアント400の表示部（図示せず）に図14に示した検索画面G₁が表示されると、ユーザは、入力部（図示せず）を用いて、検索キーとなる入力キーワード（この場合、「CHOCOA」）をキーワード入力欄B₁に入力した後、検索ボタンB₂を押下する。

【0042】

これにより、クライアント400からは、入力キーワード（「CHOCOA」）がネットワーク200を介して検索部360へ送信される。そして、入力キーワード（「CHOCOA」）が検索部360に受信されると、検索部360は、入力キーワード（「CHOCOA」）を検索キーとして図5に示した検索情報データベース350内を1レコード単位で検索し、入力キーワード（「CHOCOA」）と「キーワード」とが一致するレコードを抽出する。そして、検索部360は、抽出されたレコードを検索結果としてバッファ（図示せず）に格納する。

【0043】

検索が終了すると、検索部360は、バッファに保存されている検索結果をネットワーク200を介して、クライアント400へ送信する。この検索結果がクライアント400に受信されると、クライアント400の表示部（図示せず）に

は、図15に示した検索結果画面G₂が表示される。この検索結果画面G₂では、入力キーワード（「CHOCO A」）に対して5件分のタイトル、文書位置情報（URL）、本文の一部が表示されている。

【0044】

そして、ユーザは、入力部を用いて、検索結果画面G₂の5件から所望の1件の文書情報を指定する。これにより、クライアント400は、当該文書情報をたとえば文書格納装置100から取得するとともに、当該文書情報に関する文書位置情報を検索装置300へ送信する。この文書位置情報は、前述した動作と同様にして、文書位置情報データベース310に格納される。

【0045】

以上説明したように、実施の形態1によれば、ユーザの利用頻度がしきい値以上の文書情報を収集基点とする所定範囲の文書情報を収集し、収集された文書情報を被検索対象とするようにしたので、検索結果に利用頻度が高い文書情報が含まれる割合を高めることができ、結果として検索効率を向上させることができる。

【0046】

（実施の形態2）

さて、上述した実施の形態1では、図17に示した構成を採用してもよい。以下では、この場合を実施の形態2として図17を参照して説明する。図17に示したWEBサーバ100A、WEBサーバ110A、インターネット200A、検索サーバ300AおよびWEBブラウザ400は、図1に示した文書格納装置100、文書格納装置110、ネットワーク200、検索装置300およびクライアント400にそれぞれ対応している。また、図17に示した検索サーバ300Aにおいて、アナライザ320A、WEBロボット340Aおよび検索エンジン360Aは、図1に示した解析部320、検索情報収集部340および検索部360にそれぞれ対応している。

【0047】

以上本発明にかかる実施の形態1および2について図面を参照して詳述してきたが、具体的な構成例はこれら実施の形態1および2に限られるものではなく、

本発明の要旨を逸脱しない範囲の設計変更等があっても本発明に含まれる。

【0048】

たとえば、前述した一実施の形態においては、文書情報を検索する機能を実現するための情報検索プログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録された情報検索プログラムをコンピュータに読み込ませ、実行することにより検索を行うようにしてもよい。記録媒体には、光ディスク、フロッピーディスク、ハードディスク等の可搬型の記録媒体が含まれることはもとより、ネットワークのようにデータを一時的に記録保持するような伝送媒体も含まれる。

【0049】

【発明の効果】

以上説明したように、本発明によれば、しきい値に対応したユーザの利用頻度の情報を収集基点とする所定範囲の情報を収集し、収集された情報を被検索対象とするようにしたので、検索結果に利用頻度が高い情報が含まれる割合を高めることができ、結果として検索効率を向上させることができるという効果を奏する。

【図面の簡単な説明】

【図1】

本発明にかかる実施の形態1の構成を示すブロック図である。

【図2】

同実施の形態1の動作原理を説明する図である。

【図3】

図1に示した文書位置情報データベース310を示す図である。

【図4】

図1に示した収集基点位置情報データベース330を示す図である。

【図5】

図1に示した検索位置情報データベース350を示す図である。

【図6】

図1に示した解析部320の動作を説明するフローチャートである。

【図7】

図6に示した追加処理を説明する図である。

【図8】

同実施の形態1で用いられる一時格納テーブル T_1 を示す図である。

【図9】

同実施の形態1で用いられる一時格納テーブル T_2 を示す図である。

【図10】

同実施の形態1で用いられる一時格納テーブル T_3 を示す図である。

【図11】

同実施の形態1で用いられる一時格納テーブル T_4 を示す図である。

【図12】

図1に示した検索情報収集部340の動作を説明するフローチャートである。

【図13】

図12に示したリンク先文書情報収集処理を説明するフローチャートである。

【図14】

同実施の形態1における検索画面 G_1 の一例を示す図である。

【図15】

同実施の形態1における検索結果画面 G_2 の一例を示す図である。

【図16】

同実施の形態1における文書画面 G_3 の一例を示す図である。

【図17】

本発明にかかる実施の形態2の構成を示すブロック図である。

【符号の説明】

100、110 文書格納装置

200 ネットワーク

300 検索装置

320 解析部

340 検索情報収集部

360 検索部

特2000-031228

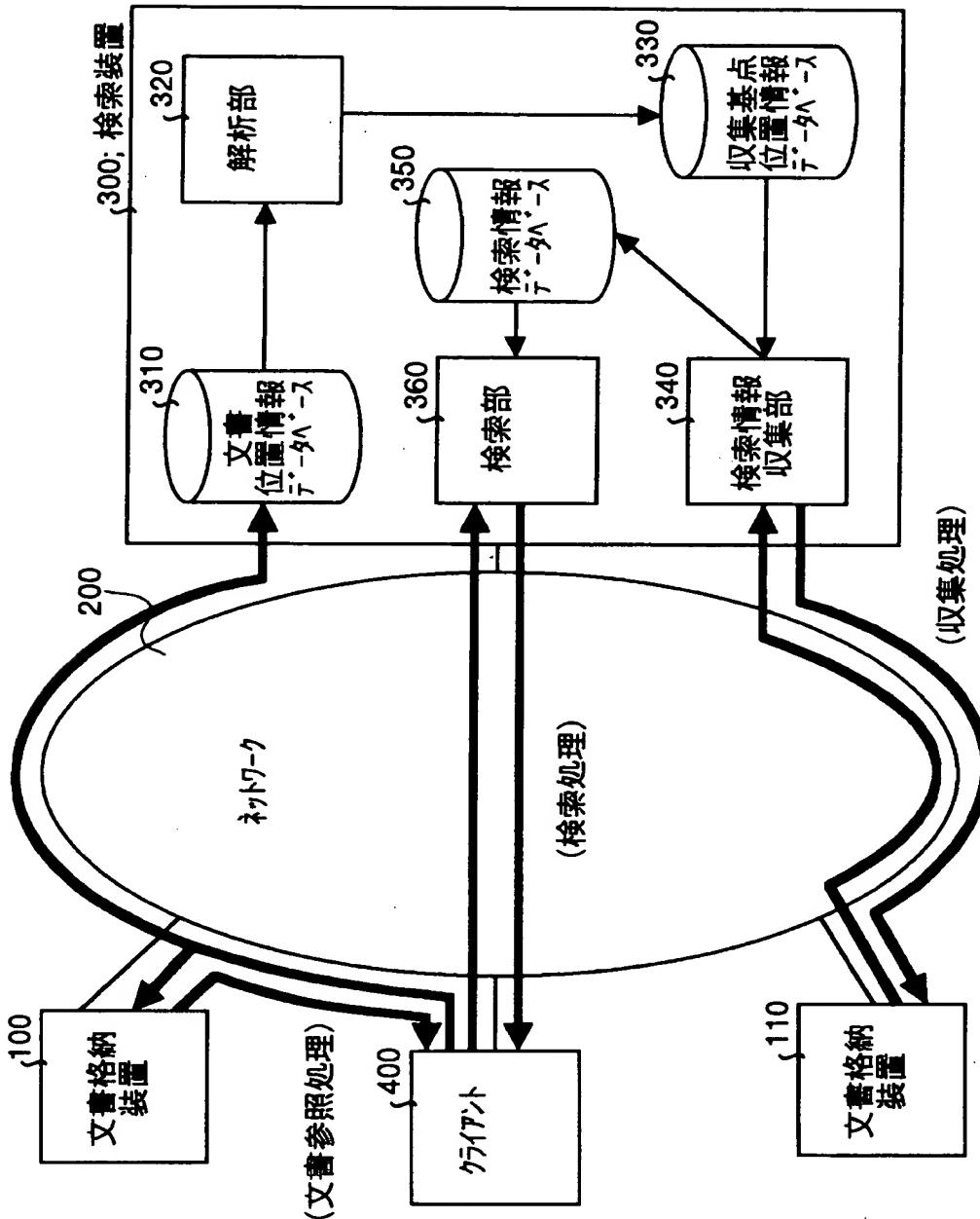
400 クライアント

【書類名】

図面

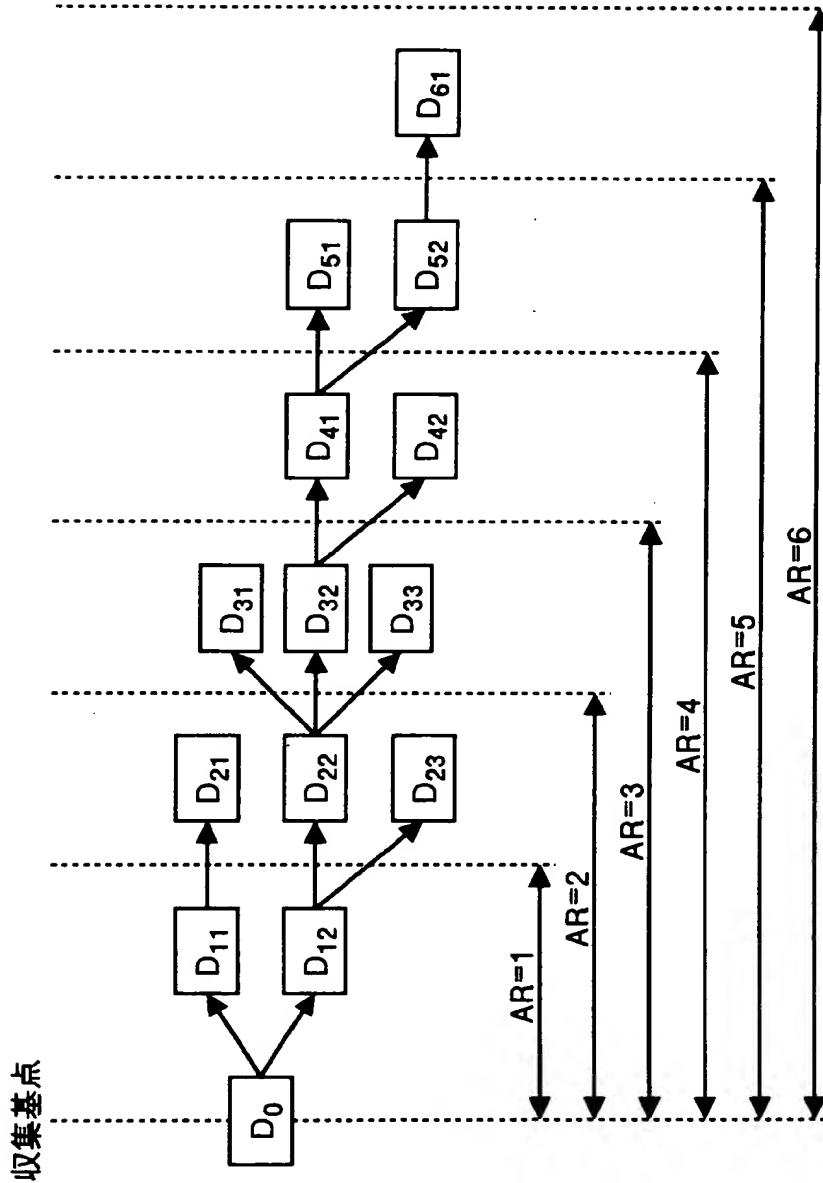
【図1】

実施の形態1の構成を示すブロック図



【図2】

実施の形態1の動作原理を説明する図



【図3】

図1に示した文書位置情報データベース310を示す図

日付	文書位置情報
1999/07/12	http://www.abcdefg.co.jp/hypertext/newinfo/
1999/07/13	http://www.is.abcdefg.co.jp/qa/qa1-10.html
	⋮
1999/07/29	http://www.is.abcdefg.co.jp/qa/qa1-10.html
	⋮
1999/08/07	http://www.ifos.se.abcdefg.co.jp/user/sekine/index.html

【図4】

図1に示した収集基点位置情報データベース330を示す図

収集基点位置情報	収集範囲
http://www.is.abcdefg.co.jp/qa/qa1-10.html	5
http://www.abcdefg.co.jp	5
http://www.ifos.se.abcdefg.co.jp/index.html	4
⋮	
⋮	
⋮	
⋮	
http://www.is.abcdefg.co.jp/qa/link.html	2

【図5】

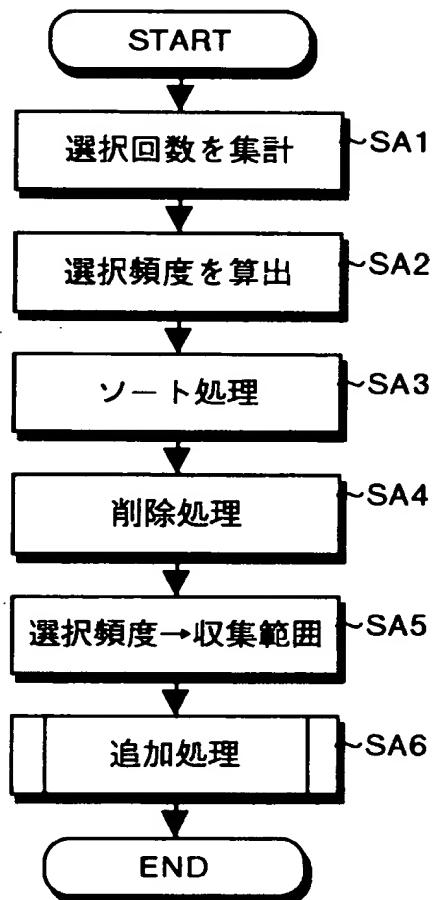
図1に示した検索情報データベース350を示す図

350

インデックス	キーワード	URL	第1表示情報	第2表示情報
1	CHOCOA	http://www.abc	Abcdefg Ja	CHOCOA1.O W
2	Windows	http://www.mic	マイクロ	製品情報
...				
10000	CHOCOA	http://www.abc	FAQs about	CHOCOA Top

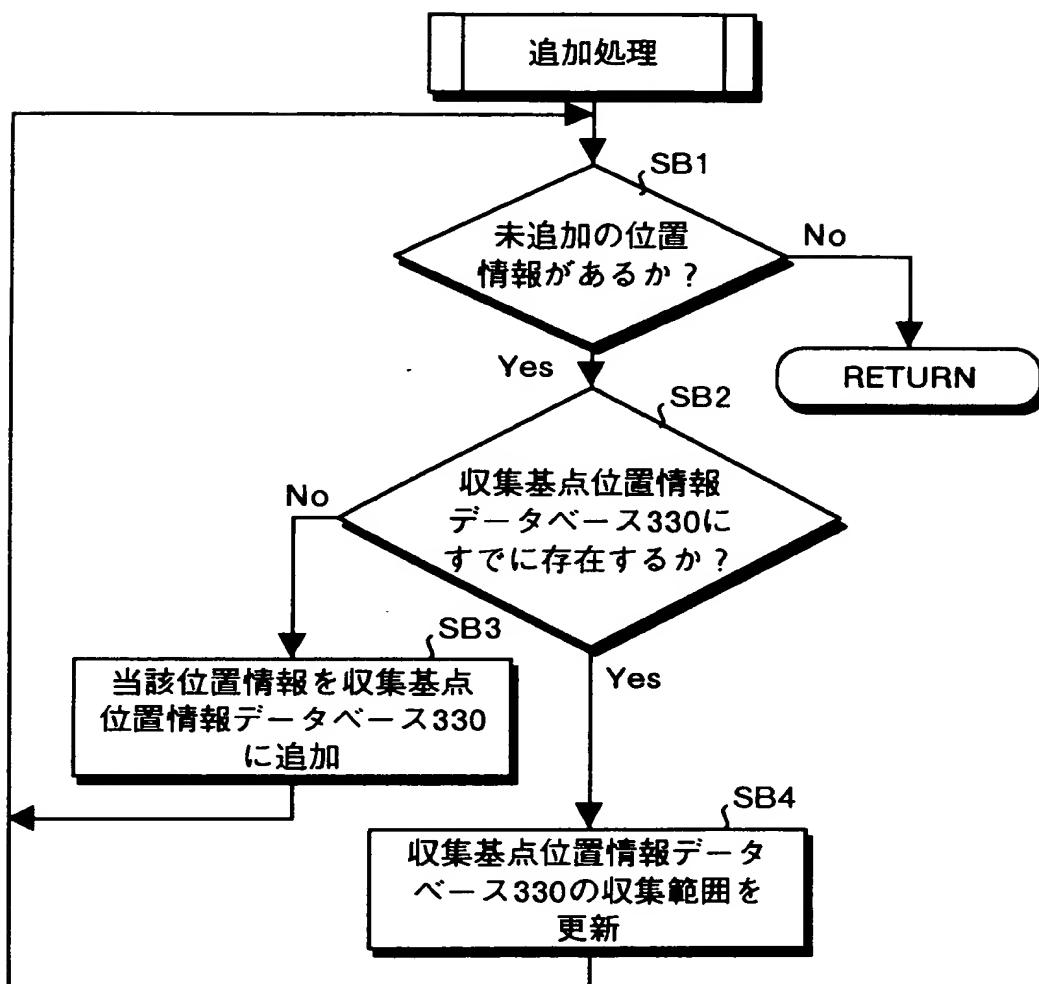
【図6】

図1に示した解析部320の動作を説明するフローチャート



【図7】

図6に示した追加処理を説明する図



【図8】

実施の形態1で用いられる一時格納テーブルT₁を示す図

文書位置情報	選択回数(回)
http://www.abcdefg.co.jp/hypertext/newinfo/	5
http://www.is.abcdefg.co.jp/qa/qa1-10.html	20
...	
http://www.ifos.se.abcdefg.co.jp/user/sekine/index.html	1
http://www.pfu.co.jp/hhk	3

【図9】

実施の形態1で用いられる一時格納テーブルT₂を示す図

文書位置情報	頻度(%)
http://www.abcdefg.co.jp/hypertext/newinfo/	10
http://www.is.abcdefg.co.jp/qa/qa1-10.html	40
...	
http://www.ifos.se.abcdefg.co.jp/user/sekine/index.html	2
http://www.pfu.co.jp/hhk	6

【図10】

実施の形態1で用いられる一時格納テーブルT₃を示す図

T ₃		
優先順位	文書位置情報	頻度(%)
1	http://www.is.abcddefg.co.jp/qa/qa1-10.html	40
2	http://www.abcddefg.co.jp/hypertext/newinfo/	10
	...	
9	http://www.pfu.co.jp/hhk	6
10	http://www.ifos.se.abcddefg.co.jp/user/sekline/index.html	2

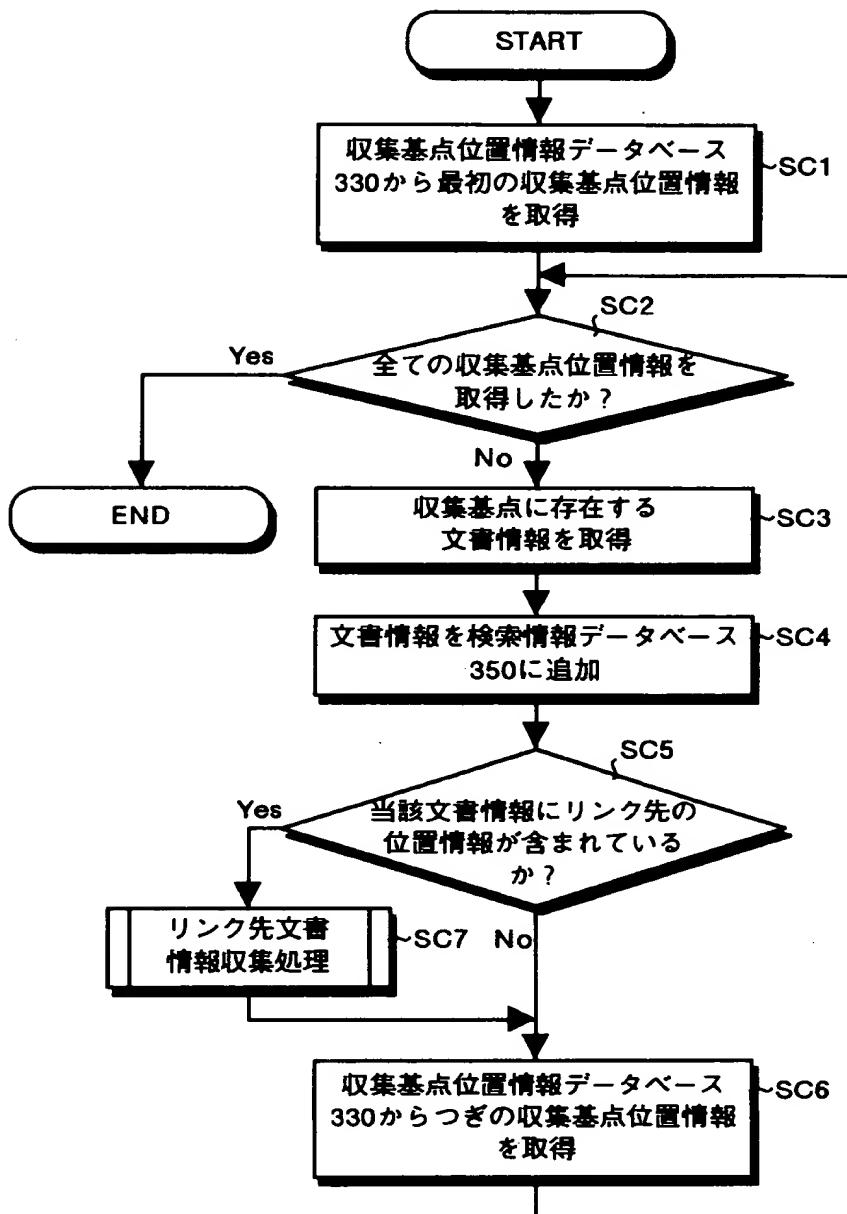
【図11】

実施の形態1で用いられる一時格納テーブルT₄を示す図

文書位置情報		T ₄
優先順位	収納範囲	
1	http://www.is.abcdefg.co.jp/qa/qa1-10.html	5
2	http://www.abcdefg.co.jp/hypertext/newinfo/	1

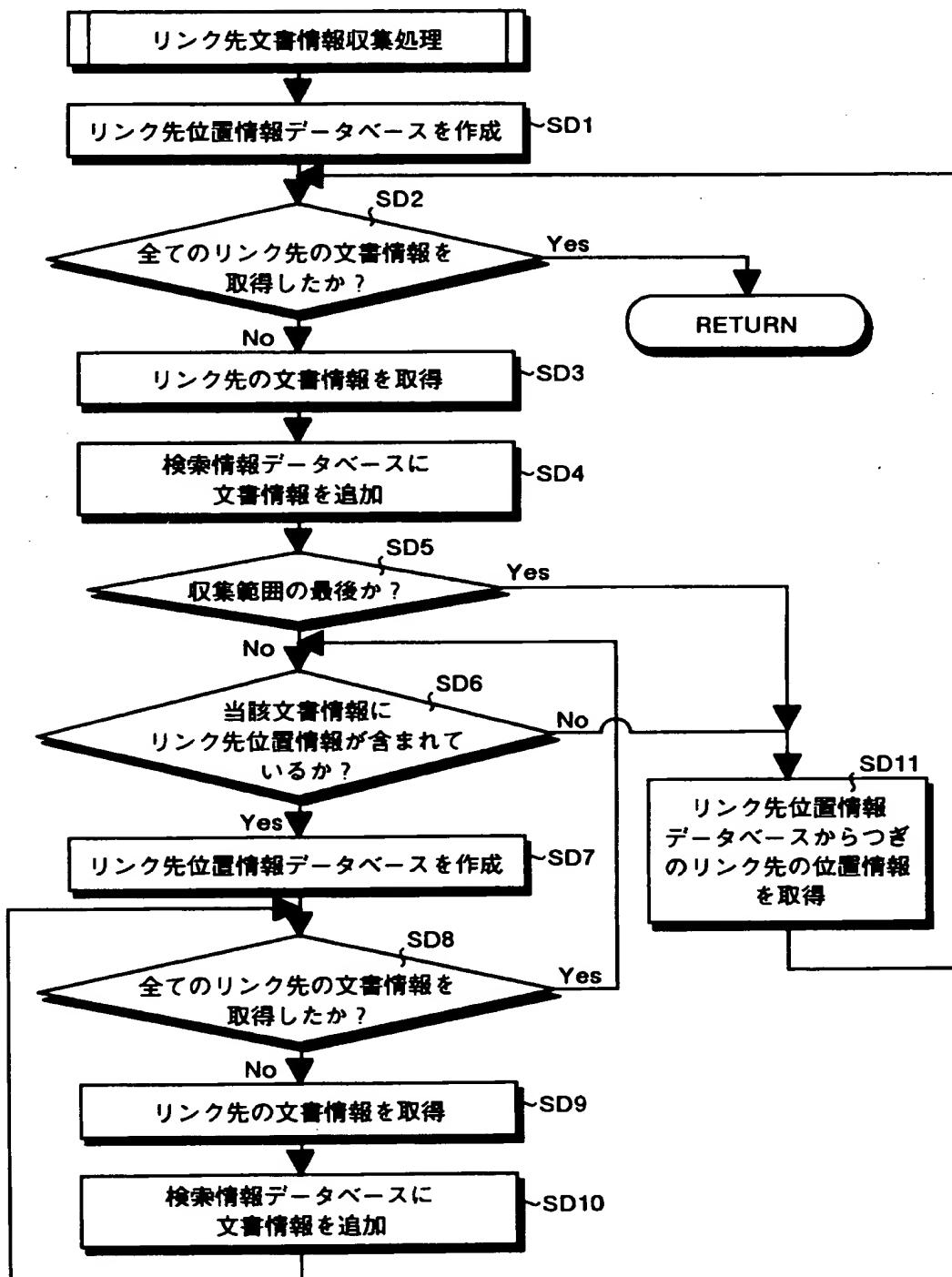
【図12】

図1に示した検索情報収集部340の動作を説明するフローチャート



【図13】

図12に示したリンク先文書情報収集処理を説明するフローチャート



【図14】

実施の形態1における検索画面G₁の一例を示す図

G₁

検索画面	B ₁	B ₂
キーワード : <input type="text" value="CHOCOA"/>	<input type="button" value="検索"/>	

【図15】

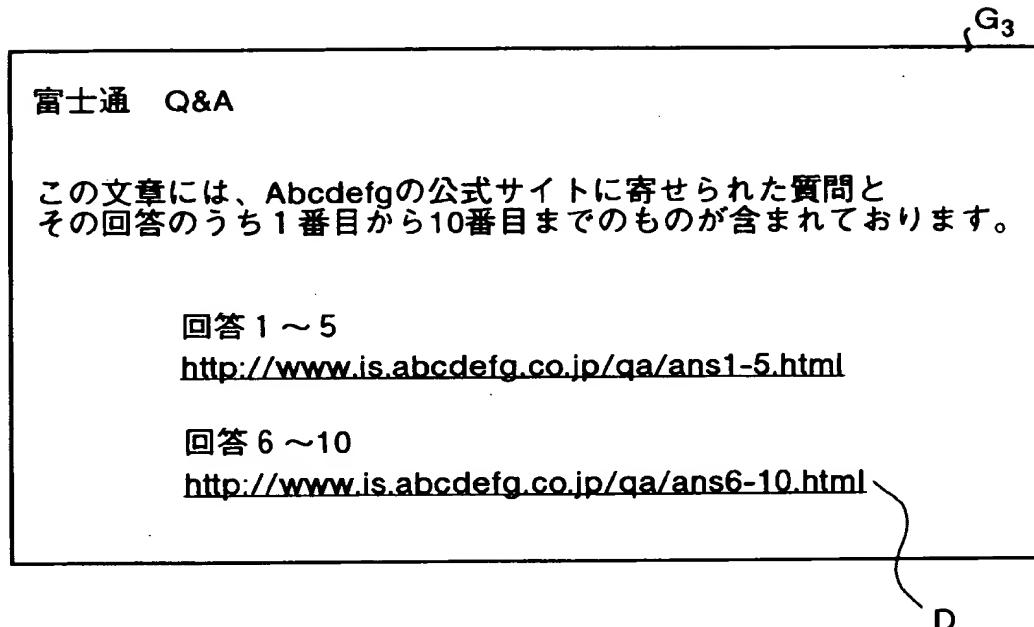
実施の形態1における検索結果画面G₂の一例を示す図

G₂

検索結果画面 5件見つかりました
1 Abcdefg Japan Homepage http://www.abcdefg.co.jp CHOCOA1.0 Windows版とフィンフィンエージェント for CHOCOAを公開
2 CHOCOA RESOURCE CENTER http://www.abcdefg.co.jp/hypertext/free/chocoa/Resource_Center.html [CHOCOA Top Page] CHOCOAリソースセンターこのページは、CHOCOAユーザ
3 CHOCOA SCRIPT FOR MACINTOSH http://www.abcdefg.co.jp/hypertext/free/chocoa/mac/CHOCOAScript/ CHOCOA 2.0で実装されているCHOCOA Scriptの機能は、CHOCOA 1.0Beta7.1
4 old News http://www.abcdefg.co.jp/hypertext/free/chocoa/oldnews.html ~CHOCOAに関する過去のトピック~ 1999年1月 2月 3月 4月 5月 6月 7月
5 FAQs about CHOCOA http://www.abcdefg.co.jp/hypertext/free/chocoa/FAQ.html [CHOCOA Top Page] CHOCOA-よく尋ねられる質問2000年対応 CHOCOAの
キーワード : <input type="text" value="CHOCOA"/> <input type="button" value="検索"/>

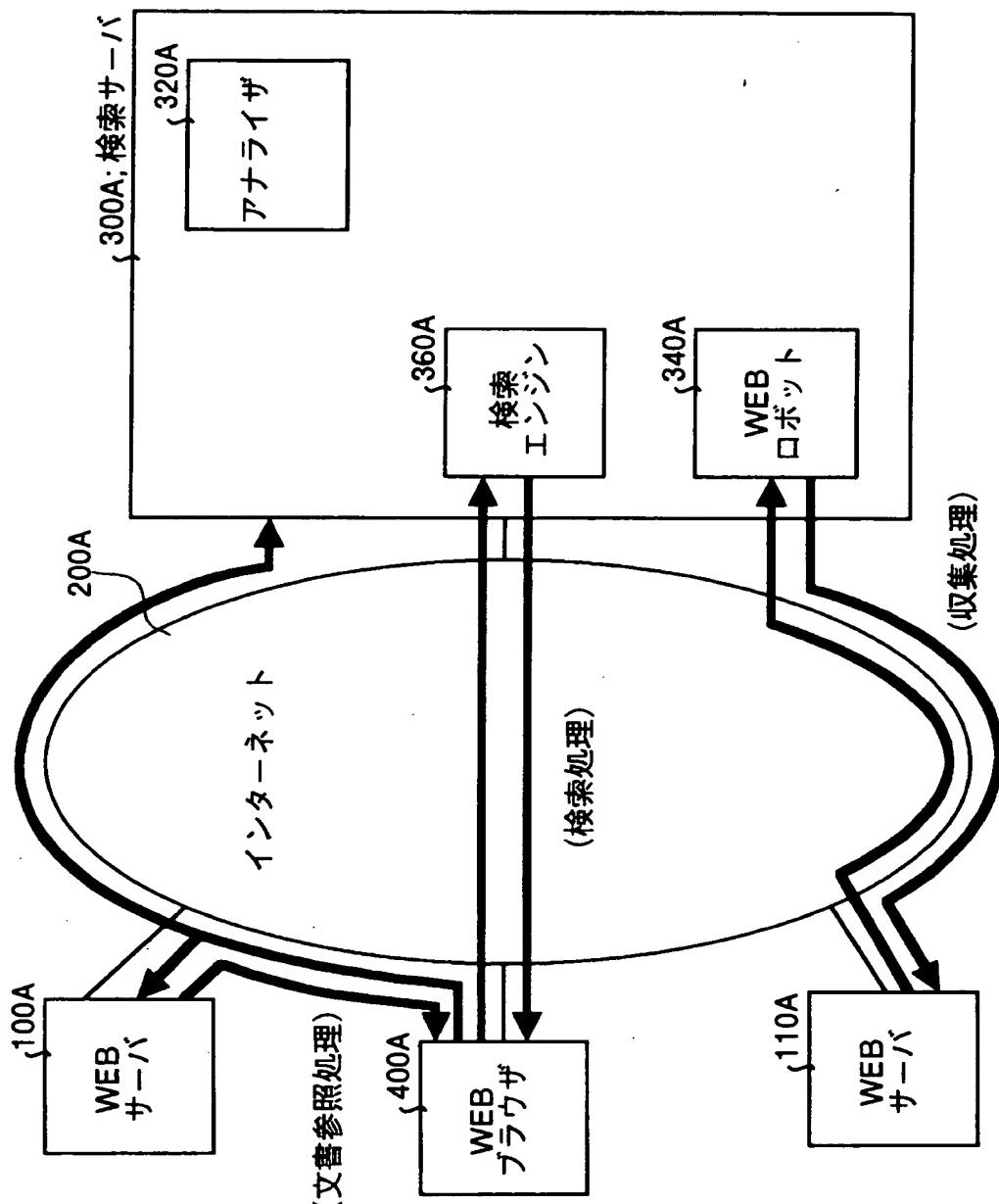
【図16】

実施の形態1における文書画面G₃の一例を示す図



【図17】

実施の形態2の構成を示すブロック図



【書類名】 要約書

【要約】

【課題】 検索効率を向上させること。

【解決手段】 ユーザにより選択された情報に関する位置情報を文書位置情報データベース310として格納する格納部と、文書位置情報データベース310内のそれぞれの位置情報の利用頻度を解析する解析部320と、しきい値以上の利用頻度の位置情報を収集基点とする所定の収集範囲に存在する情報を検索情報データベース350として収集する検索情報収集部340と、ユーザにより指定された検索条件に基づいて、検索情報データベース350から所望の情報を検索する検索部360とを備えている。

【選択図】 図1

【書類名】 手続補正書
【整理番号】 9951551
【提出日】 平成12年 2月 9日
【あて先】 特許庁長官 殿
【事件の表示】
【出願番号】 特願2000- 31228
【補正をする者】
【識別番号】 000005223
【氏名又は名称】 富士通株式会社
【代理人】
【識別番号】 100089118
【弁理士】
【氏名又は名称】 酒井 宏明
【手続補正 1】
【補正対象書類名】 明細書
【補正対象項目名】 発明の名称
【補正方法】 変更
【補正の内容】 1
【ブルーフの要否】 要

【発明の名称】 情報検索装置および情報検索プログラムを記録したコンピュータ読み取り可能な記録媒体

出願人履歴情報

識別番号 [000005223]

1. 変更年月日 1996年 3月26日

[変更理由] 住所変更

住 所 神奈川県川崎市中原区上小田中4丁目1番1号
氏 名 富士通株式会社